Figures 3A-and 3B illustrate the *HY2* locus of Arabidopsis. Figure 3A shows a map of the region of chromosome 3 containing *HY2*. Two distinct mapping populations were screened, and mapping results with molecular markers are summarized schematically, indicating that HY2 lies in a region 66 kb in length. Markers starting with the letter c are CAPS markers developed during this study. DNA sequence information for bacterial artificial chromosomes (BACs) MZB10 and F3L24 is available in GenBank/EMBL/DDBJ. The *HY2* gene structure with mutations is illustrated at the bottom. Exons are depicted as dark boxes and thick lines, which reflect coding regions and 59/39 untranslated regions, respectively. Dotted lines indicate introns. Figure 3B shows the genomic sequence of *HY2* (SEQ ID NO:32) and the deduced HY2 protein sequence (SEQ ID NO:33) from the Columbia (Col) ecotype. Uppercase letters represent exons determined by sequence analysis of HY2 cDNAs. Introns and spacer sequences are indicated with lowercase letters. The stop codon is double underlined. Mutations in hy2 alleles are shown in boldface letters. Single nucleotide polymorphisms in both Ler and Wassilewskija (Ws) ecotypes include the following: inserted T (at nucleotide 234), G364T conversion with amino acid change to Asn, and G1182A conversion (silent). Single nucleotide polymorphisms in the Ler ecotype only include the following: C515A (in intron), G884A (silent), C1145T (in intron), and G1717A (in intron). The single nucleotide polymorphism in Ws ecotype only is C1910T (silent).

a'

Delete the paragraph at page 13, line 25 through page 14, line 13, and insert the following:

Figure 5 shows an alignment of HY2 and HY2-Related Proteins. Alignment of the HY2 protein (SEQ ID NO:34) with proteins of unknown function from oxygenic photosynthetic bacteria identified by PSI BLAST. Conserved residues in 100 or 80% of the aligned sequences are depicted in the consensus sequence with uppercase or lowercase letters, respectively. Sequence similarity groups shown in the consensus sequence reflect conservation in 100% of the sequences. These are labeled as follows: 1, D 5 N; 4, R 5 K; 5, F 5 Y 5 W; and 6, L 5 I 5 V 5 M. Dark shading with white letters, gray shading with white letters, and gray shading with black letters reflect 100, 80, and 60% sequence conservation, respectively. Sequence identifiers correspond to hypothetical proteins from *Synechococcus* sp WH8020 (YCP2_SYNPY (SEQ ID NO:35) and YCP3_SYNPY (SEQ ID NO:36)), from *Prochlorococcus* (YHP2_PROMA (SEQ ID NO:37) and YHP3_PROMA (SEQ ID NO:38)), and from *Synechocystis* sp PCC 6803 gene (cyanobase locus slr0116 (SEQ ID NO:39); see http://www.kazusa.or.jp/cyano/cyano.html). Database accession numbers are AB045112 for HY2

a2

a2 cont

(DDBJ), Q02189 for YCP2_SYNPY (SWISSPROT), Q02190 for YCP3_SYNPY (SWISSPROT), CAB95700.1 for YHP2_PROMA (EMBL), CAB95701.1 for YHP3_PROMA (EMBL), and S76709 for slr0116 (Protein Information Resource). Asterisks are indicated every 20 residues.

Delete the paragraph at page 15, lines 1-27, and insert the following:

Figure 10 shows a multiple sequence alignment of the identified bilin reductases. All identified sequences were aligned using the programs CLUSTAL W and MEME. Conserved residues in 90 or 70% of the aligned sequences are depicted in the consensus sequence with uppercase or lowercase letters, respectively. Sequence similarity groups, labeled 1 (D, E), 2 (R, K), 3 (F, Y ,W), and 4 (L, I, V, M),, shown in the consensus sequence reflect conservation in >90% of the sequences. Dark shading with white letters, gray shading with white letters, and gray shading with black letters reflect 90, 70, and 50% sequence conservation, respectively. SYNY3, *Synechocystis sp* PCC6803; SYNPY, *Synechococcus sp* WH8020; SYN81, *Synechococcus sp* WH8102; PROMA, *Prochloroccocus sp* SS120; PROME, *Prochloroccocus sp* MED4; NOSPU, *Nostoc punctiforme*; ANASP, *Anabaena sp* PCC7120; ARATH, *Arabidopsis thaliana*; and HORVU, *Hordeum vulgare.* Database accession numbers are GB: AF339056 for PcyA_ANASP (SEQ ID NO:40) (CyanoBase contig 362), GB: AF339057 for PcyA_NOSPU (SEQ ID NO:41) (JGI contig 632), PIR: S76709 for PcyA_SYNY3 (SEQ ID NO:42), PcyA_SYN81 (SEQ ID NO:43) is on JGI contig 51, GB: AF352050 for PcyA_PROME (SEQ ID NO:44) (JGI contig 26), SW: Q02189 for PebA_SYNPY (SEQ ID NO:45), PIR: S31075 (fragment)/ JGI contig 72 for PebA_SYN81 (SEQ ID NO:46), EMB: CAB95700.1 for PebA_PROMA (SEQ ID NO:47), PebA_PROME (SEQ ID NO:48) is on JGI contig 26, GB: AF352049 for PebA_NOSPU (SEQ ID NO:49) (JGI contig 622), SW: Q02190 for PebB_SYNPY (SEQ ID NO:50), PebB_SYN81 (SEQ ID NO:51) is on JGI contig 72, EMB: CAB95701.1 for PebB_PROMA (SEQ ID NO:52), PebB_PROME (SEQ ID NO:53) is on JGI contig 26, GB: AF339058 for PebB_NOSPU (SEQ ID NO:54) (JGI contig 622), DDBJ: AB045112 for HY2_ARATH (SEQ ID NO:55), EMB: CAB77705.1 for RCCR_HORVU (SEQ ID NO:56), EMB: CAB16763.1 for RCCR_ARATH (SEQ ID NO:57). Asterisks indicate every tenth amino acid; dashes indicate gaps; numbers above the line indicate amino acid sequence numbering starting with number one.

a3